

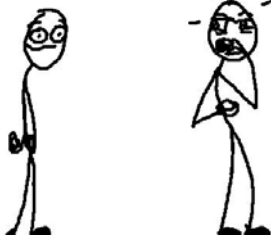
Multiple Regression II and III

REVISED

SIX and SEVEN

This worksheet relates to sections 3.2, 6.2, 12.6, 14.3, 13.6, 14.1-14.2 of the text book (Statistics for Managers 4th Edition).

Utter Revulsion.



So that we are able to have a week of revision before your mid-semester exam we have had to combine multiple regression 2 and 3. Not everything will be able to be covered so please revise tutorial and lecture material as well.

CALCULATION QUESTIONS

1. An architect believes that the relationship between the cost of constructing a multi-level car park and the area in the car park is given by the quadratic regression model:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$$

where Y is the cost per square metre in dollars and X is the area in 100 000s of square metres. He has collected the following data:

Y	X
16	2.6
19	3.4
22	4.3
26	4.5
28	5
29	6.2
30	6.8
33	7.2
36	8.4
40	9.7

The following output is for the linear regression model:

Dependent Variable: Y				
Method: Least Squares				
Date: 09/17/06 Time: 14:44				
Sample: 1 10				
Included observations: 10				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	8.993869	1.423310	6.318981	0.0002
X	3.254067	0.229916	14.15327	0.0000
R-squared	0.961597	Mean dependent var	27.90000	
Adjusted R-squared	0.956796	S.D. dependent var	7.475144	
S.E. of regression	1.553748	Akaike info criterion	3.896073	
Sum squared resid	19.31306	Schwarz criterion	3.956590	
Log likelihood	-17.48037	F-statistic	200.3150	
Durbin-Watson stat	1.274808	Prob(F-statistic)	0.000001	

The following output is for the quadratic regression model:

Dependent Variable: Y				
Method: Least Squares				
Date: 09/17/06 Time: 14:44				
Sample: 1 10				
Included observations: 10				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	4.470924	3.653321	1.223797	0.2606
X	4.948230	1.289089	3.838547	0.0064
X^2	-0.138824	0.104089	-1.333706	0.2241
R-squared	0.969378	Mean dependent var	27.90000	
Adjusted R-squared	0.960629	S.D. dependent var	7.475144	
S.E. of regression	1.483230	Akaike info criterion	3.869647	
Sum squared resid	15.39981	Schwarz criterion	3.960423	
Log likelihood	-16.34824	F-statistic	110.7969	
Durbin-Watson stat	1.637717	Prob(F-statistic)	0.000005	

(a) Estimate the linear regression model.

(b) Estimate the quadratic regression model.

- (c) Predict the cost of constructing a car park that is 200 metres by 200 metres and has 6 levels. Use both equations.

- (d) Test the significance of β_2 using 5% level of significance. What can we conclude about which estimation is better for the data?



Make sure you are very familiar with the assumptions underlying multiple linear regression - you will DEFINITELY be asked about them on your exam

2. An analyst for the US Environmental Protection Agency is studying the relationship between the speed at which a car travels (S) and the amount of petrol consumed per mile (G). It is believed that the variables are approximately related in a nonlinear way according to the equation:

$$G = \alpha_0 S^{\alpha_1} \exp(\varepsilon)$$

where G represents petrol consumption in miles per gallon, S represents the average speed of the car in miles per hour, and ε is an error term. A car is driven around a track at different speeds under carefully controlled conditions. The results are reported in the table below. Find the least squares estimates of the unknown parameters.

G	S
36	30
30	35
25	40
23	45
20	50
19	55
17	60
16	65
14	70
13	75

- (a) Transform the above equation so as to be able to estimate the regression model

- (b) Estimate the regression equation using to output below

Dependent Variable: ln(G)
Method: Least Squares
Date: 09/17/06 Time: 15:23
Sample: 1 10
Included observations: 10

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	7.216058	0.109492	65.90496	0.0000
ln(S)	-1.073084	0.027851	-38.52881	0.0000
R-squared	0.994640	Mean dependent var	3.008907	
Adjusted R-squared	0.993970	S.D. dependent var	0.328145	
S.E. of regression	0.025482	Akaike info criterion	-4.324827	
Sum squared resid	0.005195	Schwarz criterion	-4.264310	
Log likelihood	23.62414	F-statistic	1484.469	
Durbin-Watson stat	2.087762	Prob(F-statistic)	0.000000	

YOU MAY NOT GET THROUGH THE FOLLOWING QUESTIONS IN YOU SESSIONS... if that is the case please try them at home and run the answers by your leaders in your next session.

3. In mining engineering, holes are often drilled through rock using drill bits. As the drill hole gets deeper, additional rods are added to the drill bit to enable additional drilling to take place. It is expected that drilling time increases with depth. This increased time could be caused by several factors, including the increased mass of the drill rods that have been strung together. A key question relates to whether drilling is faster using dry drilling holes or wet drilling holes. Dry drilling holes involve forcing compressed air down the drill rods to flush the cuttings and drive the hammer. Wet drilling holes involve forcing water down the rods instead of air. A researcher uses the following model to predict additional drilling time based on depth and type of drilling hole (wet or dry):



$$TIME = \beta_0 + \beta_1 DUM + \beta_2 DEPTH + \varepsilon$$

where $TIME$ = time taken to drill an additional metre (in minutes)

$DEPTH$ = depth (in metres)

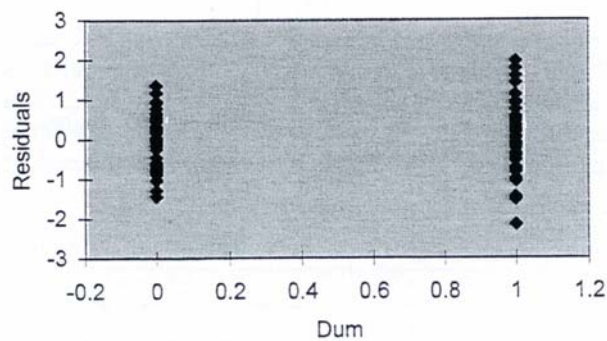
DUM = type of drilling (1 = wet, 0 = dry)

and ε is an error term. The following results are obtained from EXCEL:

<i>Regression Statistics</i>	
Multiple R	0.834
R Square	0.696
Adjusted R Square	0.6899
Standard Error	0.750
Observations	100

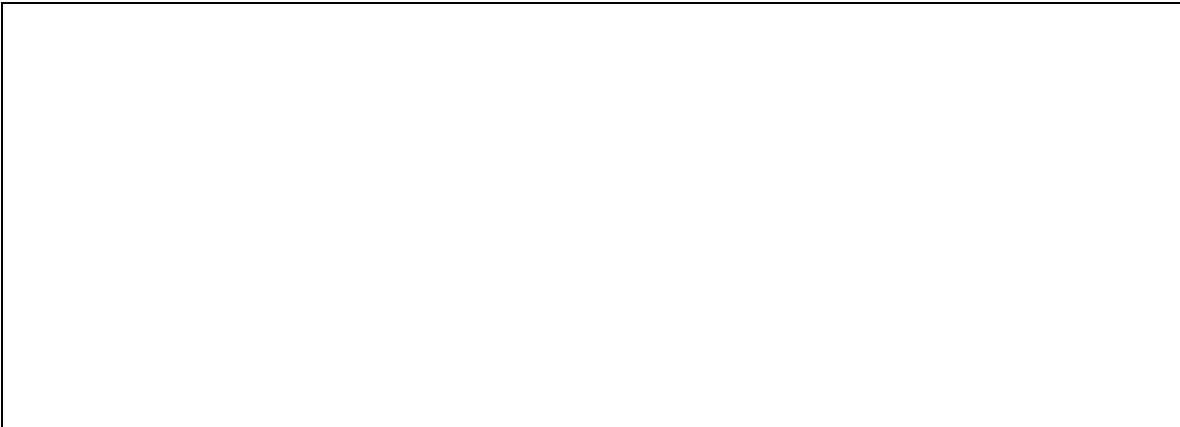
ANOVA				
	<i>Df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>
Regression	2	125.0	****	****
Residual	97	****	****	
Total	99	179.6		

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>	<i>P-value</i>
Intercept	5.90	0.17	****	****
Dum	2.11	0.15	****	****
Depth	0.005	0.001	****	****

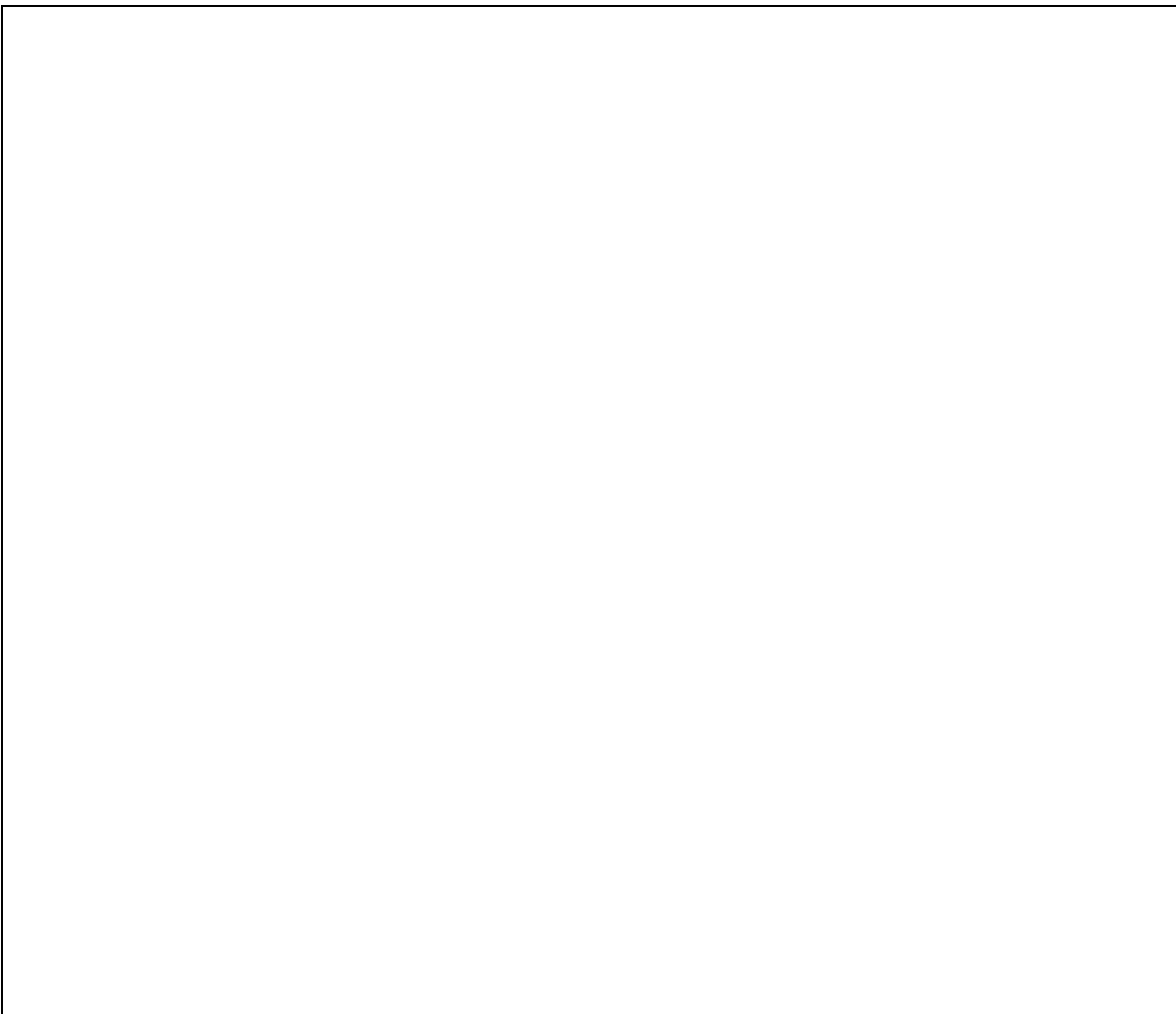


Durbin-Watson Calculations	
Sum of Squared Difference of Residuals	74.1
Sum of Squared Residuals	54.6
Durbin-Watson Statistic	1.357743871

(a) Explain the meaning of the dummy variable coefficient



(b) conduct a test to determine whether errors are positively correlated



- (c) Is there evidence of heteroskedasticity? Explain.

MULTIPLE CHOICE QUESTIONS

1. If the plot of the residuals is fan-shaped, which assumption is violated?
- (a) normality
 - (b) homoscedasticity
 - (c) independence of errors
 - (d) none – the graph should resemble a fan

Final Exam, Nov 2004

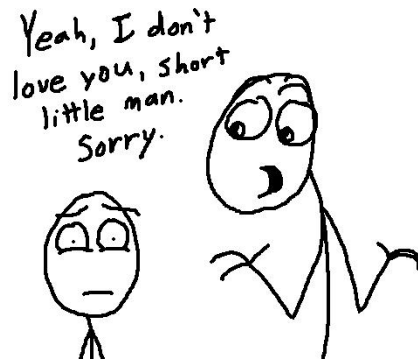
2. Problems of collinearity exist when
- (a) two or more of the explanatory variables are correlated with each other
 - (b) one of the explanatory variables is correlated with the dependent variable
 - (c) adjacent error terms are correlated with each other
 - (d) the model is not properly specified

Final Exam, Nov 2004

3. If the Durbin-Watson statistic has a value close to 4, which assumption is violated?
- (a) normality
 - (b) independence of errors
 - (c) homoscedasticity
 - (d) none of the above

Final Exam, Nov 2004

4. which of the following statements is TRUE?
- (a) the Durbin-Watson statistic can be used to check the assumption of normality.
 - (b) if the residuals in a regression analysis of time ordered data are positively correlated, the value of the Durbin-Watson statistic should be near 4.
 - (c) if the Durbin-Watson statistic takes a value near 0, then the errors are unlikely to be autocorrelated.
 - (d) if the Durbin-Watson statistic takes a value near 0 then the regression model probably violates the regression assumption of independence of errors.



notes